605100

605100

LINEAR PROGRAMMING

Robert E. Kalaba

P-950

October 1, 1956

The RAND Corporation
1700 MAIN ST. · SANTA MONICA · CALIFORNIA

Best
Available
Copy

SUMMARY

Topics include: ~~In this tutorial paper we give several~~ examples of linear programming problems, ~~formulate~~ the central mathematical problem, ~~consider various~~ methods of solution, and ~~indicate~~ lines along which further research is required.( )

LINEAR PROGRAMMING

1. Introduction. This is truly the age of "systems analysis". And yet the mathematical tools available for making such analyses are rather limited. In fact, solving a system of several hundred linear equations is about at the limit of present day computing machines. Nonlinear systems present even more imposing problems in their analysis, leaving aside questions of optimization. During the past decade, however, a new mathematical tool, linear programming, has been developed which has proved its usefulness in a wide variety of fields and holds even greater promise for the future. Here we shall present a number of significant linear programming problems, formulate the central mathematical problem, consider methods of solution, and indicate lines along which farther research is required. Our objective is to provide a general view of the subject; though the treatment is of necessity concise, references to the available literature are provided for the interested reader [2], [4], [7], [9], [12].

2. Examples. In this section we shall briefly indicate the wide range of applicability of linear programming by formulating three illustrative problems. The first two are very important in the linear programming literature, and the third, the optimal routing problem (and other related optimal telecommunication systems design problems) [8], is of particular importance to communications engineers and managers.

The Transportation Problem [1], [5], [6], [9]. An item is produced at each of m sources and consumed at each of n sinks. Source i produces a_i units, and sink j requires b_j units. The sources produce just enough to fulfill the total demands at the sinks. If the unit cost of shipment from source i to sink j is c_{ij}, find a shipping schedule which minimizes the cost of shipping.

If we let x_{ij} be the amount shipped from i to j, the mathematical problem is to minimise

$$\sum_{j=1}^{n} \sum_{i=1}^{m} c_{ij}x_{ij} ,$$

subject to the constraints

(2.1a) $$x_{ij} \geq 0 ,$$

(2.1b) $$\sum_{j=1}^{n} x_{ij} = a_i ,$$

(2.1c) $$\sum_{i=1}^{m} x_{ij} = b_j .$$

The expression to be minimised is the total cost of shipment; condition (2.1a) expresses the fact that the amount shipped from source i to sink j is nonnegative; Eq. (2.1b) indicates that the sum of the items shipped from source i to all the sinks is a_i; Eq. (2.1c) shows that the sum of the items shipped to sink j from the sources is b_j.

<u>The Nutrition Problem</u>. To produce suitable animals for market it is determined that m nutrients are required, the daily requirement of nutrient i being b_i. Various feeds are available, n in number, the amount of nutrient i in food j being a_{ij} and the unit cost of feed j being c_j. Determine a minimal cost diet.

If we let x_j be the quantity of feed j to be used daily, the problem is to minimise the cost

$$\sum_{j=1}^{n} c_j x_j ,$$

subject to the conditions

(2.2a) $$x_j \geq 0 ,$$

(2.2b) $$\sum_{j=1}^{n} a_{ij} x_j = b_i .$$

Eq. (2.2a) expresses the nonnegativity of the variables and Eq. (2.2b) is a statement of the requirement that the sum of the nutrients of type i contributed by the various feeds in the diet must equal b_i.

<u>The Optimal Routing Problem</u> [8]. A telecommunication network consists of n stations each of which may send, receive and relay messages. The stations are joined by communication channels, the capacity of the channel from i to j being c_{ij} (messages per unit time). The number of messages originated at i and destined for j per unit time is a_{ij}. Find a routing doctrine which maximizes the number of messages delivered per unit time.

If we let x_{ijk} $(i \neq j,\ i \neq k)$ be the number of messages per unit time sent over the direct link from i to j, ultimately destined for k, mathematically the problem is to maximize

$$\sum_{j=1}^{n} \sum_{i=1}^{n} x_{ijj} ,$$

subject to the conditions

(2.3a) $$x_{ijk} \geq 0 \, ,$$

(2.3b) $$\sum_{k=1}^{n} x_{ijk} \leq c_{ij} \, ,$$

(2.3c) $$\sum_{k=1}^{n} x_{ikj} - \sum_{k=1}^{n} x_{kij} \leq a_{ij} \, .$$

The number of messages sent per unit time from i to j, destined for j is x_{ijj}: thus, the sum of this quantity over all i and j ($i \neq j$) is the total number of messages delivered per unit time. The variables are nonnegative. Condition (2.3b) expresses the condition that the total number of messages sent per unit time over the channel from i to j is to be no greater than c_{ij}. Condition (2.3c) is a type of conservation law. It is an expression of the requirement that the total number of messages per unit time sent out from i, ultimately destined for j, less the total number of messages sent in per unit time to i, ultimately destined for j, is no greater than the total number of messages originated per unit time at i, ultimately destined for j (The difference between the right hand and left hand expressions is the number of messages per unit time backlogged at i, ultimately destined for j.).

3. The General Problem [3], [9], [14]. It will be observed that in each of these problems one is to determine an optimal program; that is, one is to find a set of nonnegative values for the variables (sometimes referred to as "activity levels") which optimize some objective, subject to known resource limitations and activity interdependencies. The

resources consumed and produced by an activity are assumed to be proportional to the level of activity, as is the payoff for the activity. Thus in the optimal routing problem the activities consist in sending messages ultimately destined for k from i to j. If a message ultimately destined for k is sent from i to j, a unit of capacity from i to j is consumed, a unit of demand from i to k is consumed, and a unit of demand from j to k is produced; there is no payoff since no message is delivered in carrying out this particular activity. If a message destined for j is sent from i to j, single units of capacity and demand from i to j are consumed, and one unit of payoff (a single delivered message) is produced. The resources in this problem are the capacities and demands. The conditions (2.3b) and (2.3c) are in effect statements that the amount of a resource consumed is not to be greater than the sum of the quantities exogenously provided and endogenously produced. They express the activity interdependencies and resource limitations relevant to this problem.

In general we consider that there are n activities and m resources. Each unit of activity j consumes a_{ij} units of resource i (By convention a_{ij} is positive if resource i is consumed, negative if it is produced, by activity j.). Resource i is given present exogenously in the amount b_i. Then

$$(3.1) \qquad \sum_{j=1}^{n} a_{ij}x_j \leq b_i \quad (i=1,2,\ldots,m)$$

is the general conservation law. If the list of activities is complete, then equality will hold. If the payoff per unit of activity j is c_j, the problem is to find nonnegative x_j which maximize the total payoff

$$\sum_{j=1}^{n} c_j x_j ,$$

subject to condition (3.1). This is one form of the central mathematical problem of linear programming. In essence we have given a general technique for reducing a large class of problems of economics, engineering and physics to mathematical form.

4. A Geometrical Version. From the heuristic point of view it is desirable to formulate a geometric picture of the situation. To this end we consider the following problem:

Maximize

$$z = 2x + 3y ,$$

subject to the constraints

(4.1) $$x \geq 0, \ y \geq 0 ,$$

(4.2) $$x + y \leq 3 ,$$

(4.3) $$x + 2y \leq 4 .$$

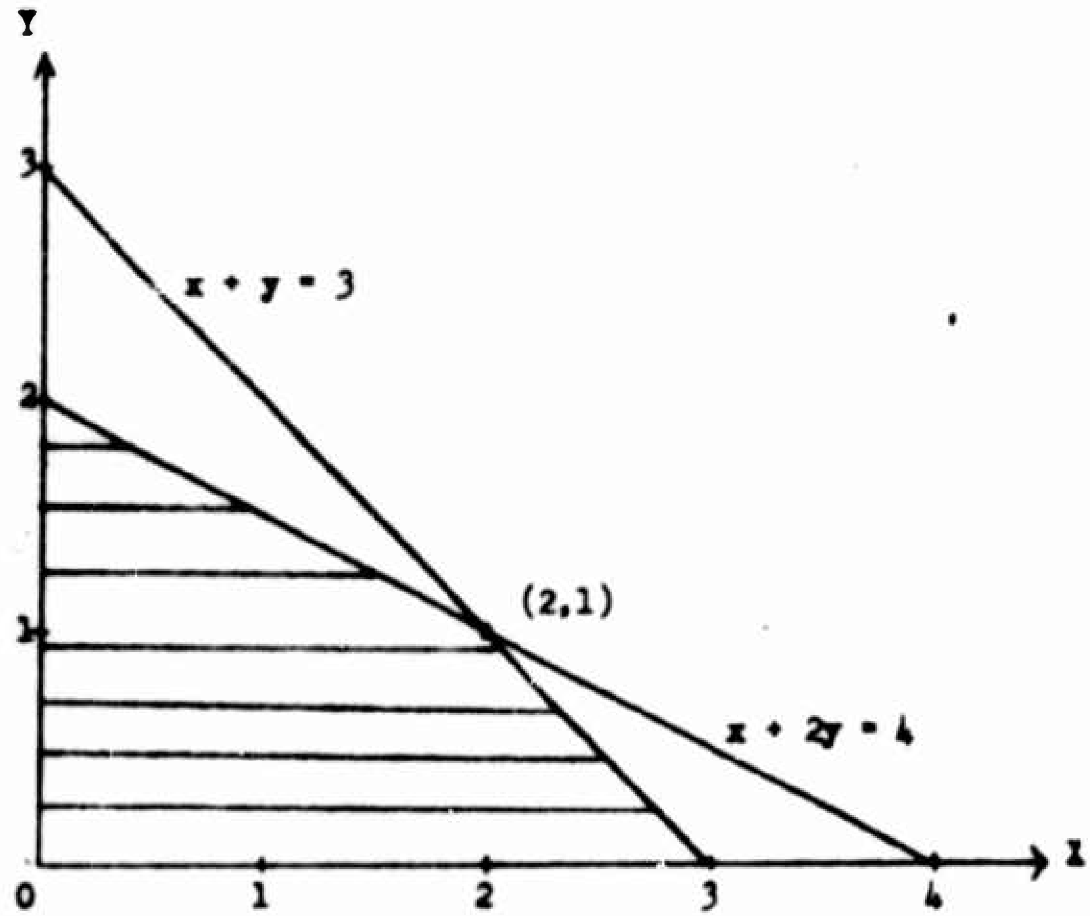

Fig. 1. THE GEOMETRY OF THE PROBLEM

We determine the set of admissible (x,y), the "feasible solutions", by noting that the nonnegativity of x and y implies that (x,y) must lie in the first quadrant. The points which satisfy condition (4.2) include the line $x + y = 3$ and all points in the half-plane containing the origin and bounded by this line. A similar remark holds for condition 4. The set of points satisfying all the conditions of the problem is the striated region of Fig. 1. In Fig. 2 we have added some equivalue lines of the objective form $2x + 3y$.

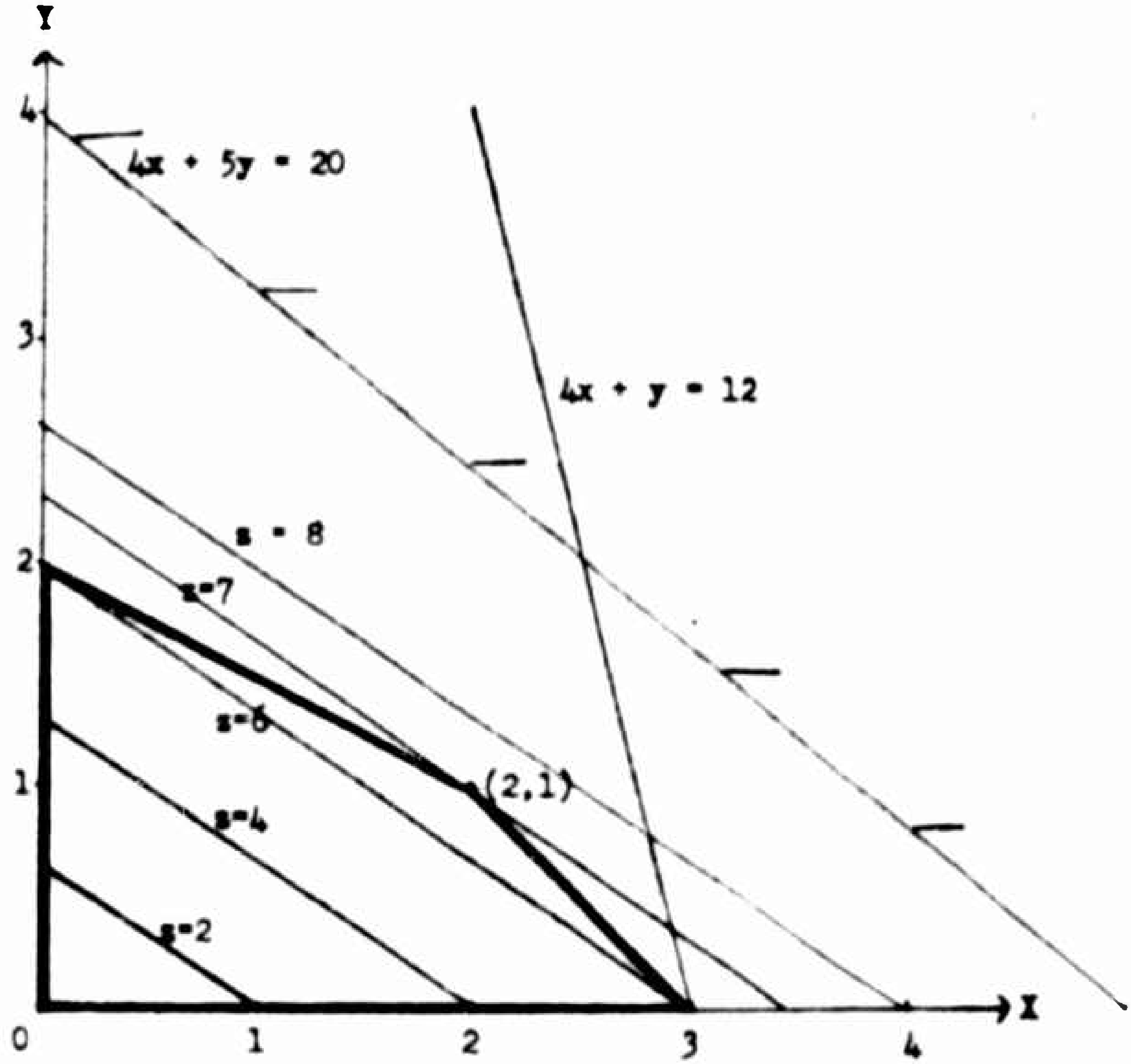

Fig. 2. DETERMINATION OF THE OPTIMAL POINT

From this figure it is readily seen that the solution lies at the point (2,1) for which z = 7. The important thing to notice is that the solution lies at a vertex of the set of feasible solutions.

If the equivalue lines of the objective form had a slope of less than -1, e.g., the slope -4 possessed by the line y = -4x + 12 shown in Fig. 2, the maximum would have occured at the vertex (3,0). If on the other hand the slope of the equivalue lines should be -1, then all points on the line segment joining (2,1) and (3,0) would be solutions -- one of the vertices could be selected -- and we would no longer have uniqueness of solution.

(It is a curious fact that non-uniqueness is the rule rather than the exception in practice.). Next we observe that if to the conditions (2), (3), and (4) we add the condition $5y + 4x \geq 20$, then there are no feasible solutions, for there are no points lying in the straited region of Fig. 1 which satisfy this additional condition. In general we may say that if a solution exists, it can be chosen at a vertex.

It is possible to program an analogue computer so that a point in the admissible region will be driven toward an optimal vertex, the point being kept in the region by using large negative feedbacks whenever it reaches a boundary [13].

Lastly we leave it to the reader to examine the following problem graphically:

Maximise

$$x + y ,$$

subject to the constraints

$$x \geq 0, \ y \geq 0 ,$$

$$y - x \leq 2 ,$$

$$y - x \geq -2 .$$

It will readily be seen that the objective form is unbounded; that is, it can be made larger than any preassigned value.

Thus we have seen that linear programming problems may have no solution (The constraints are inconsistent or the objective form is unbounded), one solution, or infinitely many solutions; no other cases arise. For problems involving two or three variables and only a few conditions one may carry through the investigations graphically. For a problem involving hundreds of variables and thousands of conditions, such

an approach is manifestly inadequate. How may such problems be treated?

5. <u>General Method of Solution</u> [3], [9]. Though other methods have been proposed [10], without a doubt the most successful technique for treating general linear programming problems is the simplex algorithm of George B. Dantzig. (A simplex is the generalization to n dimensions of the triangle of two dimensions.) A number of modifications exist, and various specialisations for particular problems have been made, but the essence of the method is the following:

The problem is taken as given in the form:

Maximise

$$z = c_1x_1 + c_2x_2 + \cdots + c_nx_n ,$$

subject to the constraints

$$x_1 \geq 0,\ x_2 \geq 0,\ \ldots,\ x_n \geq 0 , \tag{5.1}$$

$$\left\{\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{12}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \qquad\qquad \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m . \end{array}\right. \tag{5.2}$$

This form is really quite general. If a given problem calls for minimisation of the objective form, as does the transportation problem, we may simply maximise the negative of the given objective form. If the variables are

not nonnegative, we may in effect make them that way by introducing two nonnegative variables for each such variable and using the fact that any real number may be written as the difference of two nonnegative numbers

$$x_i = x_i' - x_i'' .$$

In addition, an inequality constraint may be made into an equality constraint by adding a nonnegative slack variable. Thus

$$a_1x_1 + a_2x_2 + \dots + a_nx_n \leq b$$

may be written

$$a_1x_1 + a_2x_2 + \dots + a_nx_n + y = b ,$$

where y is nonnegative.

It is possible to demonstrate that

(a) If any feasible solution to the equality system (5.2) exists, then one exists involving only as many variables as there are equations, the remaining variables being set equal to zero.

(b) If feasible solutions exist and these possess a finite minimum for the objective form z, then an optimizing solution exists involving only as many variables as there are equations, the remaining variables being set equal to zero.

The simplex method takes advantage of these theorems by first showing how to obtain a feasible solution involving no more than m non-zero variables, should such a solution exist. A simple test is employed to determine whether or not the solution is optimal. If it is, the algorithm

ceases, for a solution has been found. If it is not, it is shown how a new variable may be added to the basic set and one of the variables in the basic set eliminated, in such a way that the value of the objective form is increased. Since there are only a finite number of bases, the algorithm will terminate after a finite number of steps.

An example will illustrate the essential points of the algorithm. Consider the problem of section 4 in the form for which z is to be maximized:

$$x,\ y,\ u,\ v \geq 0 \tag{5.3}$$

$$x + y + u \quad = 3 \tag{5.4}$$

$$\underline{x + 2y + \quad v = 4} \tag{5.5}$$

$$2x + 3y \quad = z \tag{5.6}$$

According to the theorems stated above, if the objective form has a finite maximum -- which we shall assume -- then at most two non-zero variables will appear in the solution. Though in large problems the determination of an initial basic feasible solution may be difficult, here we see that we could choose, for example, x = 3, v = 1, and y and u zero. We eliminate x from Eq. (5.5) by subtracting Eq. (5.4) from it, so that x appears in just one condition and v in the other. Then x and v are eliminated from the objective form by multiplying Eq. (5.4) by 2 and subtracting from Eq. (5.6). These operations yield

$$
\begin{array}{llr}
(5.7) & x + y + u = 3 & \\
(5.8) & \underline{\underline{ y - u + v = 1}} & \\
(5.9) & y - 2u = s - 6\ . &
\end{array}
$$

Thus s has the value 6. But this is not optimum, for we can increase s by increasing y. What prevents y from being increased indefinitely? We note that as y is increased, v and x have to be decreased to preserve equality in conditions (5.7) and (5.8). Here u is kept at zero value, for increasing-u would decrease s. This can be continued until y = 1, for further increase would drive v negative. Thus y is increased to 1, x decreased to 2, v decreased to 0, u remains at 0, and s is increased to 7. Now we have a basic feasible solution in which x and y are non-zero, u and v are zero.

Upon eliminating y from Eqs. (5.7) and (5.9) we obtain

$$
\begin{array}{ll}
(5.10) & x + 2u - v = 2 \\
(5.11) & \underline{\underline{ y - u + v = 1}} \\
(5.12) & - u - v = s - 7
\end{array}
$$

But now we see that this solution is optimal, for the left side of Eq. (9) has its maximum value when u = v = 0.

Though a number of subtleties have gone unmentioned, the main trend of thought is illustrated by the above example. The simplex method lends itself well to computation by high-speed electronic digital computers. Problems involving 255 conditions and any number of variables may be

solved using a code devised by W. Orchard-Hays for the IBM 704 computer [11]. Problems involving 50 conditions take just a few minutes to solve.

6. Conclusion. From the examples of preceding sections it is evident that linear programming has been applied to problems in a number of fields. To those already mentioned we may add design of optical systems and beam structures, scheduling of aircraft, and oil refining. The list is growing quite rapidly. In addition, various connections between linear programming and the theory of games are known.

Much, however, remains to be done. As more and more complicated models are considered, problems involving ever increasing numbers of conditions and variables are generated. An optimal routing problem involving 30 stations, for example, gives rise to a linear programming problem with 1,740 conditions and 25,230 variables. Though general problems of this sort may be beyond present day computation capabilities, research making use of the special structure of various problems offers great possibilities. Transportation problems involving 3,200 equations and 600,000 variables are currently being programmed. Improved computing machines and computation techniques, together with better understanding of construction of models should make linear programming a tool of ever increasing importance in the future.

REFERENCES

1. A. Boldyreff, "Determination of the Maximal Steady State Flow of Traffic Through a Railroad Network", Journal of the Operations Research Society of America, Vol. 3 (1955), pp. 443-465.

2. A. Charnes, W. Cooper, A. Henderson, An Introduction to Linear Programming, New York (1953).

3. G. Dantzig, A. Orden, and P. Wolfe, "The Generalized Simplex Method for Minimizing a Linear Form under Linear Inequality Restraints", Pacific Journal of Mathematics, Vol. 5 (1955), pp. 183-195.

4. K. Eiseman, "Linear Programming", Quarterly of Applied Mathematics, Vol. 8 (1955), pp. 209-232.

5. L. Ford, D. Fulkerson, "A Simple Algorithm for Finding Maximal Network Flows and an Application to the Hitchcock Problem", The RAND Corporation Paper P-743 (1955).

6. F. L. Hitchcock, "The Distribution of a Product from Several Sources to Numerous Localities", Journal of Mathematics and Physics, Vol. 20 (1941), pp. 224-230.

7. A. J. Hoffman, "Linear Programming", Applied Mechanics Reviews, Vol. 9 (1956), pp. 185-187.

8. R. E. Kalaba and M. L. Juncosa, "Communication Networks I. Optimal Design and Utilization", The RAND Corporation Research Memorandum RM-1687 (1956) (To appear in Management Science).

9. T. C. Koopmans (ed.), Activity Analysis of Production and Allocation, New York (1951).

10. C. E. Lemke, "The Dual Method of Solving the Linear Programming Problem", Naval Research Logistics Quarterly, Vol. 1 (1954), pp. 36-47.

11. W. Orchard-Hays, "The Evolution of Linear Programming Techniques", The RAND Corporation Paper P-900 (1956).

12. Proceedings of the Second Symposium in Linear Programming, Vols. I and II, Washington, D. C. (1955).

13. I. B. Pyne, "Linear Programming on an Electronic Analogue Computer", Communications and Electronics, May 1956, pp. 139-142.

14. S. Vajda, The Theory of Games and Linear Programming, New York (1956).